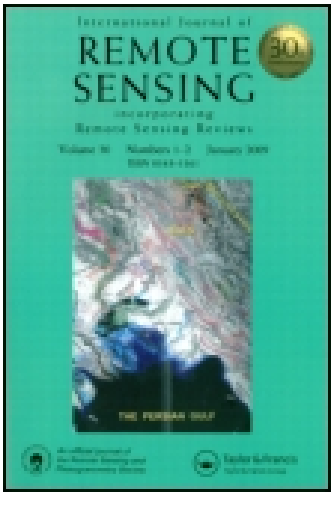

International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Applicability of an automatic surface detection approach to micro-pulse photon-counting lidar altimetry data: implications for canopy height retrieval from future ICESat-2 data

Mahsa S. Moussavi[a], Waleed Abdalati[a], Ted Scambos[ab] & Amy Neuenschwander[c]

[a] Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA

[b] National Snow and Ice Data Center (NSIDC), University of Colorado, Boulder, CO, USA

[c] Applied Research Laboratories, University of Texas at Austin, Austin, TX, USA

Published online: 21 Jul 2014.

To cite this article: Mahsa S. Moussavi, Waleed Abdalati, Ted Scambos & Amy Neuenschwander (2014) Applicability of an automatic surface detection approach to micro-pulse photon-counting lidar altimetry data: implications for canopy height retrieval from future ICESat-2 data, International Journal of Remote Sensing, 35:13, 5263-5279, DOI: 10.1080/01431161.2014.939780

To link to this article: http://dx.doi.org/10.1080/01431161.2014.939780

International Journal of Remote Sensing, 2014
Vol. 35, No. 13, 5263–5279, http://dx.doi.org/10.1080/01431161.2014.939780

Taylor & Francis
Taylor & Francis Group

Applicability of an automatic surface detection approach to micro-pulse photon-counting lidar altimetry data: implications for canopy height retrieval from future ICESat-2 data

Mahsa S. Moussavi[a]*, Waleed Abdalati[a], Ted Scambos[a,b], and Amy Neuenschwander[c]

[a]*Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA;* [b]*National Snow and Ice Data Center (NSIDC), University of Colorado, Boulder, CO, USA;* [c]*Applied Research Laboratories, University of Texas at Austin, Austin, TX, USA*

(*Received 30 August 2013; accepted 13 June 2014*)

We develop and validate an automated approach to determine canopy height, an important metric for global biomass assessments, from micro-pulse photon-counting lidar data collected over forested ecosystems. Such a lidar system is planned to be launched aboard the National Aeronautics and Space Administration's follow-on Ice, Cloud and land Elevation Satellite mission (ICESat-2) in 2017. For algorithm development purposes in preparation for the mission, the ICESat-2 project team produced simulated ICESat-2 data sets from airborne observations of a commercial micro-pulse lidar instrument (developed by Sigma Space Corporation) over two forests in the eastern USA. The technique derived in this article is based on a multi-step mathematical and statistical signal extraction process which is applied to the simulated ICESat-2 data set. First, ground and canopy surfaces are approximately extracted using the statistical information derived from the histogram of elevations for accumulated photons in 100 footprints. Second, a signal probability metric is generated to help identify the location of ground, canopy-top, and volume-scattered photons. According to the signal probability metric, the ground surface is recovered by locating the lowermost high-photon density clusters in each simulated ICESat-2 footprint. Thereafter, canopy surface is retrieved by finding the elevation at which the 95th percentile of the above-ground photons exists. The remaining noise is reduced by cubic spline interpolation in an iterative manner. We validate the results of the analysis against the full-resolution airborne photon-counting lidar data, digital terrain models (DTMs), and canopy height models (CHMs) for the study areas. With ground surface residuals ranging from 0.2 to 0.5 m and canopy height residuals ranging from 1.6 to 2.2 m, our results indicate that the algorithm performs very well over forested ecosystems of canopy closure of as much as 80%. Given the method's success in the challenging case of canopy height determination, it is readily applicable to retrieval of land ice and sea ice surfaces from micro-pulse lidar altimeter data. These results will advance data processing and analysis methods to help maximize the ability of the ICESat-2 mission to meet its science objectives.

1. Introduction

The Ice, Cloud, and land Elevation Satellite (ICESat) has returned unprecedented data on elevation changes over ice sheets, sea ice freeboard heights, cloud vertical distribution, and vegetation canopy height (e.g. Harding 2005; Kurtz et al. 2008; Zwally et al. 2005; Schutz et al. 2005; Shuman et al. 2006; Lefsky et al. 2005; Thomas et al. 2008; Yi, Zwally, and Sun 2005; Magruder et al. 2007; Fricker 2005; Kwok et al. 2007; Pang et al.

*Corresponding author. Email: Mahsa.Moussavi@colorado.edu

2008; Zwally et al. 2008; Popescu et al. 2011; Rosette, North, and Suárez 2008). With ICESat's successful demonstration of spaceborne lidar technology for cryosphere and ecosystem applications, NASA (National Aeronautics and Space Administration) is developing a follow-on mission, ICESat-2, scheduled to be launched around 2017. The ICESat-2 mission is expected to extend the time series of ICESat observations in order to fully characterize the trends in ice sheet and sea ice changes and detect the differences in these trends over time. It will carry a micro-pulse photon-counting lidar operating at 532 nm wavelength, named ATLAS (Advanced Topographic Laser Altimeter System) (Abdalati et al. 2010). While measurement of land ice elevation, sea ice freeboard, and changes in these variables is the primary science objective, estimation of terrestrial biomass and carbon storage through assessments of vegetation canopy heights is an additional planned capability of the ICESat-2 mission (Abdalati et al. 2010). The mission's ecosystem science requirement states that 'ICESat-2 shall produce elevation measurements that enable independent determination of global vegetation height, with a groundtrack spacing of less than 2 km over a 2-year period (Level-1 science requirements and mission success criteria available at http://icesat.gsfc.nasa.gov/icesat2/mission_overview.php)'.

As NASA moves forward with the development of the ICESat-2's ATLAS instrument, the degree to which the current instrument design will yield data beneficial to the vegetation and ecosystem science community is not yet known. This uncertainty originates from the application of a fundamentally different altimetry technology in ICESat-2 as compared with the original ICESat mission. The use of a lower per-pulse laser energy (25–100 μJ) and photon-counting detection mechanism in the proposed ATLAS instrument configuration poses certain challenges for canopy height determination from future ICESat-2 data. Further challenges arise from the low reflectivity of soil (0.3) and vegetation (0.1) and low atmospheric transmission at 532 nm. Depending upon vegetation type, crown density, atmospheric conditions, and solar elevation, the number of detected signal photons reflected from the canopy-top and the underlying ground surface is expected to range between 0 and 10 for each laser shot. Additionally, the presence of photon events associated with solar background noise and atmospheric scattering (above and within the canopy and below the ground surface) further complicates the detection of signal returns from vegetation and ground under canopy. The low signal levels, combined with the substantial background noise, will ultimately result in a very low signal-to-noise ratio (SNR) per footprint, introducing challenges to automatic retrieval of vegetation height. Therefore, sophisticated techniques are required to accurately extract the ground and canopy surfaces to enable canopy height determination from future ICESat-2 data. Several Science Definition Team (SDT) members have been working to prototype automated algorithms for canopy height determination. As part of ICESat-2's SDT efforts and in direct support of the mission, we examine airborne micro-pulse lidar data, sub-sampled to the level expected for ICESat-2, in order to determine how best to extract vegetation canopy height. Though not a definitive statement of what ICESat-2 capabilities will be, this work develops and validates an automated approach for vegetation height retrieval from micro-pulse photon-counting data that would be applicable to ICESat-2, and provides important insights into the challenges to be expected and methods to overcome them.

2. ICESat-2 measurement concept

The architecture of ICESat-2's ATLAS instrument will differ from the Geoscience Laser Altimeter System (GLAS) that flew on ICESat. In contrast to GLAS's single-beam, high-

energy, waveform-digitizing infrared lidar system, ATLAS will be a multi-beam, low-energy (25–100 μJ), photon-counting laser altimeter system operating at 532 nm (Yu et al. 2010). The multi-beam design enables measurement of cross-track slopes on a per-orbit basis, which were previously determined via several repeat-track and crossover analyses (Abdalati et al. 2010). The currently planned six-beam system has a 3 × 2 configuration, wider in cross-track direction, so that the footprints sweep out three pairs of tracks, consisting of strong and weak beams, that are separated by 3.3 km. Each laser pulse will illuminate a 10 m-diameter spot on the ground every 70 cm (e.g. Abdalati et al. 2010; Yu et al. 2010).

The main differences between the ICESat-1 and -2 missions are the high- *versus* low-SNR system design and analogue *versus* digital detection mechanisms. In general, high-SNR systems, such as GLAS, favour simplified designs over weight, size, and power consumption, and do not use available photons efficiently (Cossio et al. 2010). ICESat-2's lidar, however, will rely on emission of much lower energy pulses at a higher repetition rate (10 KHz), which will enable a denser spatial sampling (Abdalati et al. 2010). The aggregate pulse energy for 100 ATLAS strong-beam laser shots over a GLAS-equivalent 70 m footprint is 10 mJ, which is about half that of GLAS.

The ATLAS instrument will record single photon events which are returned from a full distribution of surfaces illuminated by the laser pulse. With only a few detected photons per laser fire, accumulation of photon returns from hundreds of laser shots in the along-track direction is required to reliably recover a target's vertical structure (Degnan 2002). The principal challenge in processing photon-counting data is the existence of noise photons due to ambient light and atmospheric scattering.

3. Collection of airborne micro-pulse lidar data over forested ecosystems

The ICESat-2 project team created ATLAS-like data from two airborne lidar campaigns collected by a commercial micro-pulse sensor developed by Sigma Space Corporation. During these campaigns in October 2009, the airborne micro-pulse lidar was flown over the Pine Barren regions of Silas Little and Cedar Bridge in New Jersey and the Smithsonian Environmental Research Center (SERC) forest located in Maryland. The data are available from the NASA Goddard ICESat-2 website (http://icesat.gsfc.nasa.gov/icesat2/).

The Pine Barrens is a heavily forested, flat, and sandy ecosystem whose flora composition is largely determined by fire frequency (Collins and Anderson 1994). Pitch pine (*Pinus rigida*) and shortleaf pine (*Pinus echinata*) are the most abundant trees here. A variety of oaks grow among the pines, including black oak (*Quercus velutina*), white oak (*Quercus alba*), post oak (*Quercus stellata*), chestnut (*Quercus prinus*), scarlet oak (*Quercus coccinea*), and blackjack oak (*Quercus marilandica*) (Collins and Anderson 1994). Here, the moderately dense canopy cover (70–80%) allows sunlight to penetrate and reach the forest floor, resulting in a well-developed and richly diversified understorey. In this article, the flightlines over the Pine Barrens are represented as Cedar-2 and Cedar-4.

The closed canopy SERC forests (with 33,500 trees of 84 species) are mostly characterized by their hardwood species. This site is mainly dominated by upland forest but also comprises floodplain forests. The tree species in the upland forest include, but are not limited to, the 'tulip poplar', several oaks (*Quercus* spp.), beech (*Fagus grandifolia*), and several hickories (*Carya* spp). The forest is also characterized by the mid-canopy of red maple (*Acer rubrum*) and sour gum (*Nyssa sylvatica*), and an understorey composed

of American hornbeam (*Carpinus caroliniana*), spicebush (*Lindera benzoin*), and paw-paw (*Asimina triloba*). The tall floodplain forest (as high as 40 m) is largely dominated by ash (*Fraxinus spp*), sycamore (*Platanus occidentalis*), and American elm (*Ulmus Americana*) (Parker 2012). The data acquired by a Sigma Space Corporation (Lanham, MD, USA) photon-counting lidar were collected along five tracks over the SERC forests during leaf-on conditions with canopy closure around 95%. Data over three of these tracks (represented in this article as SERC-1, SERC-3, and SERC-5) were used to create the simulated ICESat-2 data.

Selection of these two forested ecosystems, the Pine Barrens and SERC, as study areas provides an opportunity to assess the performance of the algorithm over moderately dense to dense canopy covers.

4. Simulated ICESat-2 data

To assess the performance of the automatic surface detection algorithm proposed in this article, we used a simulated ICESat-2 data set that was created from the Sigma Space lidar data over two forested ecosystems (as discussed in the previous section). The Sigma Space lidar instrument employed beam-scanning (100 beams), micro-pulse photon-counting technology operating at 532 nm, and a high repetition rate of 100 KHz (Barbieri et al. 2009). Though employing similar technology, ICESat-2's ATLAS instrument will have only six beams operating at fixed angles and a lower repetition rate of 10 kHz. As a result of the design differences between the ATLAS instrument and the Sigma Space lidar, the collected signal photon density was substantially higher than expected for ICESat-2. Hence, to produce ATLAS-like data, the airborne lidar data were down-sampled both geometrically and radiometrically (Herzfeld et al. 2013; Barbieri et al. 2009, 2010).

The data were first edited by Sigma Space such that many photons above the canopy surface and below the terrain surface were eliminated to produce a signal-only data set for further simulation. Then, the ICESat-2 project team resampled the photon returns in the signal-only data set at a point spacing and photon density expected for ICESat-2 (Barbieri et al. 2009). A brief description of the methodology used to produce ICESat-2 simulated data is provided below.

To emulate the spatial distribution of the future ICESat-2 data, 10 m-diameter footprints were specified every 70 cm along the aircraft ground track (Barbieri et al. 2009). For a given ICESat-2 footprint, the number of signal photons was calculated from a Poisson distribution with the mean parameter equal to the expected signal estimates, based on the ATLAS instrument performance model for vegetated surfaces (Martino 2010). The mean number of signal photons per shot for our study sites was modelled at 1.93, representing the high-energy beam case in ICESat-2's six-beam configuration (Martino 2010). Within each simulated footprint, the location from which to select a return photon was then determined by a randomly generated angle and radial distance. The radial distance, r, was calculated from a Gaussian-weighted distribution (2-sigma diameter = 10 m) and the angle α was calculated using a uniform distribution (Figure 1). The closest photon to this location was selected within a 1 m buffer; however, if no photons were found within this buffer size then none were selected for that footprint (Barbieri et al. 2009). The choice of buffer size (also referred to as cap size) mainly depended on data density and computer time required for photon selection (see Herzfeld et al. (2013) for detailed information). For each footprint location, random selection of photons was repeated until the desired number of photons (as specified by a Poisson distribution) was reached.

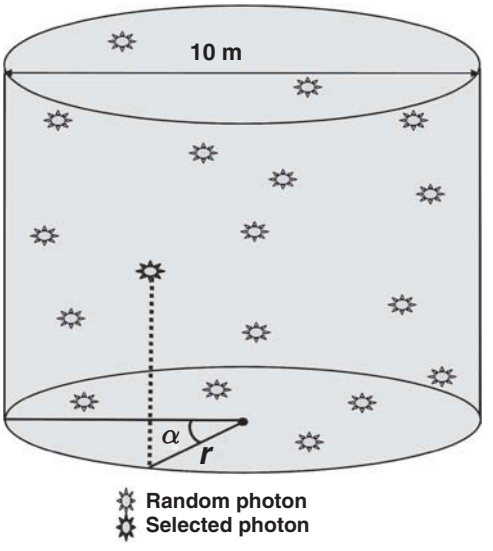

Figure 1. Schematic diagram illustrating the random selection of signal photons when performing spatial down-sampling of the full-resolution airborne photon-counting data. r and α are randomly generated angle and radial distance, respectively, from which the location of a photon return is selected within each ICESat-2 simulated footprint.

To further simulate ATLAS data, random noise events were added back to the data set (Barbieri et al. 2010). The major source of background noise in photon-counting systems is scattered solar radiation of the same wavelength as the laser output (Degnan 2002). To mimic solar background noise typical of night-time, daylight (clear sky), and daylight (high humidity) conditions, noise rates of 0.5, 2, and 5 MHz, respectively, were added to the simulated ATLAS data sets (Martino 2010). A noise rate of N MHz represents an average of N recorded noise photons for each microsecond of time corresponding to 150 m of vertical range. Due to the stochastic nature of atmospheric scattering, the number of noise photon returns at the detector was simulated using a Poisson distribution (e.g. Yang et al. 2011; Liu et al. 2006):

$$P_k(\lambda) = \frac{\lambda^k \exp(-\lambda)}{k!}, \tag{1}$$

where $P_k(\lambda)$ is the probability of k noise photons arriving at the detector when the expected number of photons is λ depending upon the noise scenario, the Poisson parameter is calculated as

$$\lambda = R_{bin} \times \frac{2\dot{n}}{c}, \tag{2}$$

where R_{bin} is the range bin width, $\dot{n}$ is the noise rate, and c is the speed of light.

It should be noted that in reality, background noise spatially varies in the along-track direction mainly as a function of surface reflectance, local atmospheric conditions, solar incidence angle, and vegetation structure (Degnan 2002). For this reason, noise photon generation was adjusted by the relative shot density of the signal within each simulated footprint. Addition of noise in this step could over-represent noise within the canopy, due

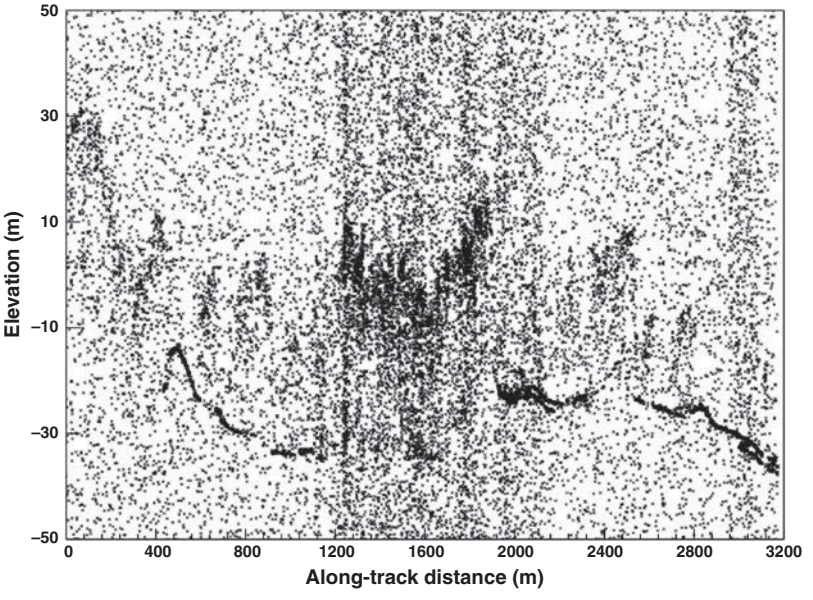

Figure 2. Example of simulated ICESat-2 data over the SERC-5 flightline (5 MHz of background noise was added to simulate humid daytime acquisition). Noise photons can be seen above and within the canopy and below the ground surface. Also evident is ground occlusion (low photon return from the ground) due to dense canopy closure.

to the fact that in the preprocessing step, only noise photons from below the ground surface and above the canopy-top were deleted. Figure 2 presents an example of simulated ATLAS data for a flightline over the SERC forest.

The ATLAS-like data sets used in this study were developed for algorithm development and assessment purposes. These were not intended to exactly duplicate the performance of the ATLAS instrument and therefore they could not directly be used to derive conclusions about the potential capabilities of ICESat-2 with regard to canopy height determination. However, they allow us to develop insight into the challenging issue of recovering ground and canopy-top signals amid noisy micro-pulse photon-counting data, which is of great utility for the ICESat-2 mission

5. Methodology

In order to determine canopy height from simulated lidar data sets, ground and canopy-top surfaces should be accurately recovered. Calculating the differences between these two surfaces will then result in canopy height estimates. Several challenges exist in detecting signal photons in the noisy photon-counting data. The main challenge arises from the presence of noise photons above and within the canopy and below the ground surface, along with the associated spatio-temporal variability of noise levels. Moreover, low SNR from the underlying terrain – especially in the case of dense forest cover – prohibits robust extraction of the ground surface and therefore impacts the accuracy of height retrieval. The objective of this research is to demonstrate a methodology that can automatically extract signal photons based on mathematical and statistical techniques. We achieve this goal through the following steps.

First, in order to roughly filter out noise events in the simulated ICESat-2 data, photon returns from 100 laser shots (equivalent to a 70 m GLAS footprint) in the along-track direction are accumulated to subsequently generate a histogram of elevations. Here, we used an elevation bin size of 1 m and elevation range of 100 m to build the histogram. The histogram h(z) is smoothed by the central moving average method, formulated as

$$\tilde{h}(z_0,2q+1) = \frac{1}{2q+1}\sum_{k=-q}^{q} h(z_j), \tag{3}$$

where $\tilde{h}(z)$ is the smoothed histogram, z is elevation, and q is the number of points around each $z_0 \in \{z_{-q}, z_q\}$ over which averaging is performed (z_{-q} and z_q are the limits of the averaging interval). The simulated data are then filtered according to the two ends of the histogram with a three-standard deviation (3σ) cut-off, removing the highest and lowest noise events in the point cloud.

To approximate the location of signal photons, we calculate and analyse the cumulative elevation distribution for 100 footprints. The inflection point of the cumulative elevation curve corresponds to the elevation at which the highest density of returned photons exists. In other words, the elevation interval within which the curve's inflection point falls represents the elevation range where a transition from low–high and high–low photon density occurs for the corresponding 70 m distance. We selected a 40 m elevation interval around the inflection point to avoid over-filtering at this stage of the analysis. Photon events were classified as signal if they occurred within the 40 m elevation buffer. Extracted signal photons in this step could be those returning from either the canopy-top, the ground surface, or those reflected from within the canopy. Some background noise photons are also misclassified as signal, but this is not of significance given that only an approximation of signal locations is desired at this point.

To help identify signal photons more accurately, we generate a signal probability metric that represents the normalized photon density around each photon in a rectangle in the XZ plane (X is the along-track dimension and Z is the elevation dimension). The signal probability vector ($\boldsymbol{\Gamma}$)for all the photons in the point cloud ($\mathbf{G}$) is populated thus:

$$\forall \boldsymbol{\Psi}_i(x_i,z_i) \in \mathbf{G}, 1 \le i \le k; \mathbf{D}_i = \{\boldsymbol{\Psi}(x,z) || x - x_i| \le a \& |z - z_i| \le b\}$$
$$\gamma_i = \frac{\mathrm{n}(\mathbf{D}_i)}{\max(\mathrm{n}(\mathbf{D}))}, \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_k \end{bmatrix}, \tag{4}$$

where $\mathbf{D}_i$ defines a cloud of individual photons ($\boldsymbol{\Psi}$), each at coordinates (x, z) around the centre photon $\boldsymbol{\Psi_i}$ at (x_i, z_i), from the first to the kth photon; a and b are the X and Z dimensions of the rectangle surrounding each photon $\boldsymbol{\Psi_i}$; γ_i is the normalized density of photons in each cloud $\mathbf{D}_i$, defined as the count of photons in each cloud ($\mathrm{n}(\mathbf{D}_i)$) divided by the maximum number of photons in any cloud ($\max(\mathrm{n}(\mathbf{D}))$); and $\boldsymbol{\Gamma}$ is the probability vector of all such normalized densities. We chose $a = 35$ m to take advantage of the horizontal trend in ground and canopy surfaces, and $b = 2$ m to avoid mixed signals from ground and tree crown surfaces.

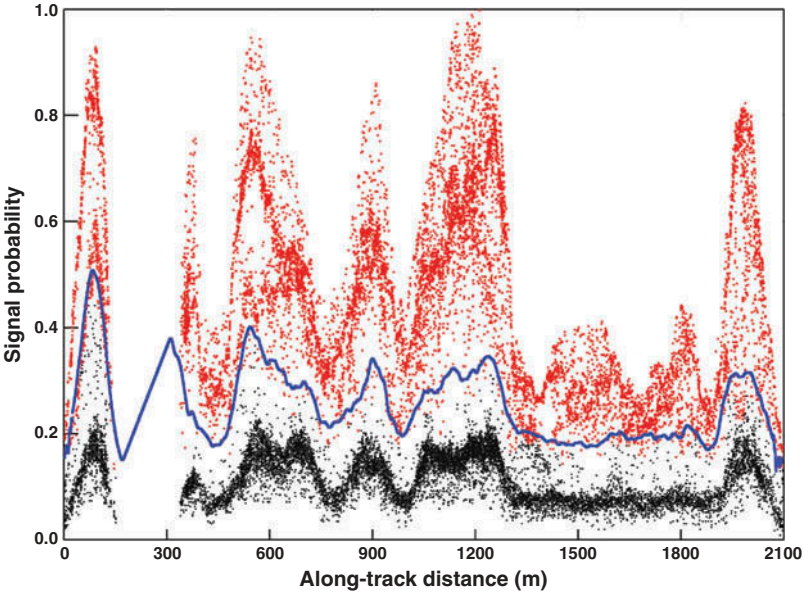

Figure 3. Signal probability metric generated for Cedar-4 at Pine Barrens (5 MHz noise rate). This metric represents the normalized photon density around each photon in a rectangle in the *XZ* plane. The red points correspond to high-photon density clusters in the data; black points represent noise photons to be removed; the blue line represents a spline fit to the lowermost high-density photons/points.

Points/photons with low signal probability(γ) correspond well with the background noise photons (Figure 3); hence, to increase the SNR, these are removed from the point cloud. For this purpose, a spline is fitted to the points of lowest signal probability. Photons that lie within 1–3 standard deviations of the spline, depending upon the noise scenario, are classified as noise and subsequently removed from the data. The filtered lidar data now mainly consist of signal photons reflected from the ground surface/tree crowns and also dense clusters of volume-scattered photons.

Using the filtered data, we recover the ground surface by finding the lowermost photons in each 10 m-diameter footprint. To eliminate the outliers and reduce the effect of misclassified noise photons, we use an iterative process based on cubic spline interpolation. If photons lie within three standard deviations of the spline fit they are classified as ground returns, otherwise they are removed. This process is repeated three times to ensure as much noise is removed as possible while still retaining enough data.

Once a ground surface is detected, we determine the canopy surface by finding the elevation at which 95% of above-ground photons exist in individual 10 m footprints. (To maintain consistency between algorithm development efforts, ICESat-2's SDT proposed to define the canopy-top as where the 95th percentile of above-ground photons exists.) To eliminate the remaining noise, the same filtering technique as in the ground location step is performed. Finally, by applying the aforementioned sequence of processing steps and differencing the ground and canopy surfaces, we derive canopy height estimates from the simulated ICESat-2 data.

6. Results and validation

To investigate the performance of the proposed algorithm, we selected simulated ATLAS data sets of high laser energy and various noise levels. We then compared our ground and canopy height estimates to those derived from the full-resolution photon-counting lidar data sets, a digital terrain model (DTM), and a canopy height model (CHM) at both SERC and the Pine Barrens forests. The DTMs and CHMs were previously generated from independent airborne discrete-return lidar data acquired in October 2008 and provided to the ICESat-2 SDT for algorithm development and assessment purposes. It should be noted that the statistics reported here do not necessarily reflect the absolute errors in canopy height estimation. Several studies suggest that airborne lidars are likely to underestimate tree height (e.g. Morsdorf et al. 2008; Zimble et al. 2003; Hancock et al. 2011). However the only way to test the true performance of the algorithm and the micro-pulse instrument would be ground surveys, which were not possible in this work.

6.1. *Algorithm performance assessment using full-resolution photon-counting data*

In this section, the retrieved ground and canopy surfaces are validated against the reference surfaces derived from the full-resolution photon-counting lidar data. Table 1 provides a summary of mean residual and root-mean-squared (RMS) errors of retrieved ground elevations and canopy heights over each flight line. The residuals indicate how well the 25 m along-track segments extracted from the reference data and the results match for 1–3 km-long flight lines at SERC and the Pine Barrens. The reported canopy height errors correspond to the final tree height errors and not the canopy-top elevation. Three data acquisition scenarios were assessed, including night-time, daytime with clear skies, and daytime with high humidity conditions. At the Pine Barrens, the ground surface retrieval errors range from 0.2 to 0.50 m, whereas canopy height residuals fall between 1.6 and 2.2 m for different noise scenarios. For the closed canopy deciduous forests of SERC, the accuracy in determining the ground surface varies from 0.8 to 2.3 m, and canopy height estimates have errors in the order of 3.9–6.6 m. In general, the residual errors worsen with increasing background noise, which is mainly due to solar illumination in daytime acquisitions. The methodology yields better accuracies for Pine Barrens than SERC. This is mainly because the coniferous forests of Pine Barrens (lower canopy cover ~80%) allow more laser light to penetrate through the canopy

Table 1. Mean and RMS residual errors of recovered surfaces, assessed against the full-resolution photon-counting data (order of subtraction: reference-retrieved).

	0.5 MHz noise (night)				2 MHz noise (clear day)				5 MHz noise (humid day)			
	Ground surface (m)		Canopy height (m)		Ground surface (m)		Canopy height (m)		Ground surface (m)		Canopy height (m)	
Flight line	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS
Cedar-2	−0.74	0.50	0.49	1.63	−0.72	0.48	−0.07	1.93	−0.67	0.39	−0.49	2.02
Cedar-4	−0.15	0.21	−0.76	2.04	−0.16	0.22	−0.48	2.24	−0.18	0.45	−0.06	1.84
SERC-1	−0.13	0.81	0.70	3.98	0.20	0.82	0.06	4.16	0.18	1.22	0.71	5.86
SERC-3	0.15	1.35	−1.31	4.20	0.03	1.08	−0.41	4.26	−0.38	2.32	−0.88	4.60
SERC-5	0.26	0.94	−2.50	5.30	0.22	1.31	−2.02	6.43	0.24	1.56	−0.93	6.66

and hit the underlying ground surface. This results in a stronger return signal from the ground and subsequently better canopy height retrieval accuracy. The strength of this technique is its success in detecting small changes in topography, since it relies on using local neighbourhood information in two dimensions rather than aggregating signals over long distances.

Figures 4 and 5 provide two examples of retrieved surfaces using this technique *versus* the reference surfaces, superimposed on the simulated ICESat-2 data. These figures indicate that the proposed technique can reliably recover top-of-canopy and ground surfaces despite the low (high) signal (noise) levels.

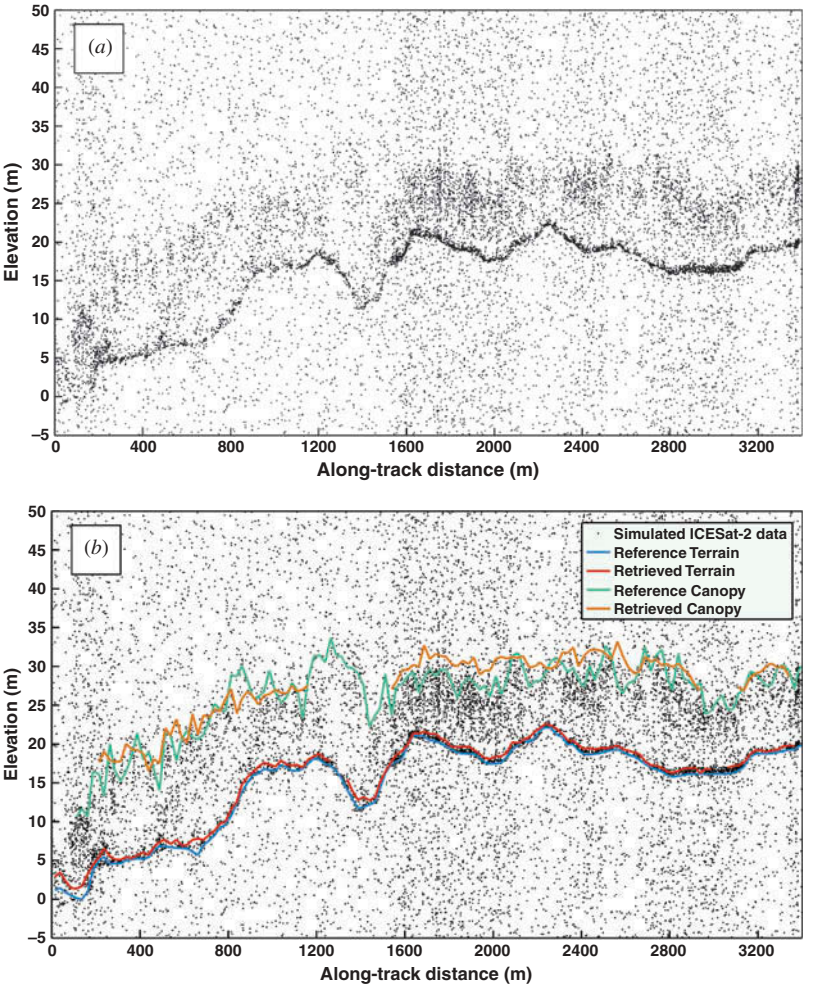

Figure 4. (*a*) Simulated ICESat-2 data over the Cedar-2 flight line with a background noise rate of 5 MHz; (*b*) Retrieved terrain and canopy surfaces over the Cedar-2 flight line overlaid on simulated ICESat-2 data and compared against reference surfaces derived from full-resolution photon-counting data. Here, canopy height is retrieved with an average error value of −0.49 m and accuracy of 2.02 m.

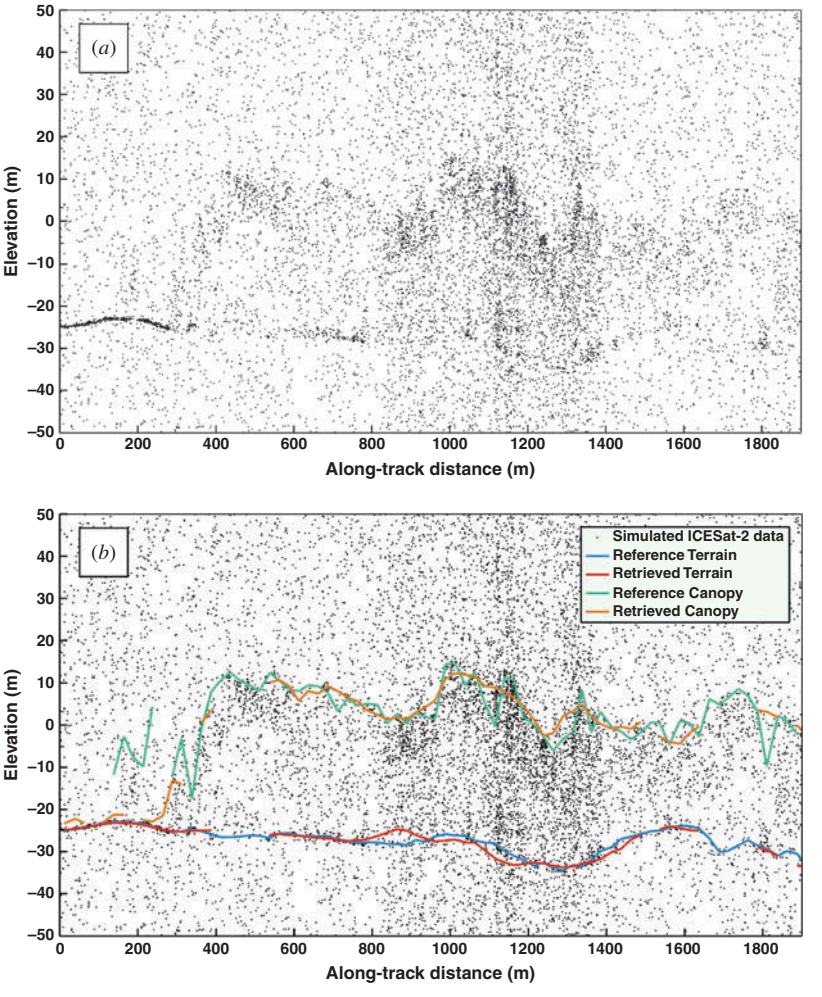

Figure 5. (*a*) Simulated ICESat-2 data over the SERC-1 flight line with a background noise rate of 5 MHz; (*b*) Retrieved terrain and canopy surfaces over the SERC-1 flight line overlaid on simulated ICESat-2 data and compared against reference surfaces derived from full-resolution photon-counting data. Here, canopy height is retrieved with an average error value of 0.71 m and accuracy of 5.86 m.

6.2. Algorithm performance assessment using DTM and CHM

To further assess the performance of the algorithm, the derived surfaces from the simulated data were evaluated against the independent discrete-return airborne lidar-derived DTMs/CHMs for the study areas. Table 2 provides a summary of the mean and RMS residuals of the retrieved surfaces for all flight lines over the Pine Barrens and SERC.

At the Pine Barrens, the ground surface RMS residuals range from 0.3 to 0.5 m, and canopy height estimates indicate residuals in the order of 3.1–3.7 m. At SERC, where the canopy cover is significantly denser than that of the Pine Barrens, the ground and canopy surfaces were resolved to an average accuracy of 1.9 and 6.0 m, respectively.

Table 2. Mean and RMS residual errors of recovered surfaces, assessed against DTM/CHM for the study area (order of subtraction: reference-retrieved).

	0.5 MHz noise (night)				2 MHz noise (clear day)				5 MHz noise (humid day)			
	Ground surface (m)		Canopy height (m)		Ground surface (m)		Canopy height (m)		Ground surface (m)		Canopy height (m)	
Flight line	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS	Mean	RMS
Cedar-2	−0.74	0.37	−4.48	3.54	−0.71	0.37	−4.77	3.66	−0.68	0.43	−4.79	3.74
Cedar-4	−0.29	0.30	−4.01	3.21	−0.30	0.30	−3.80	3.14	−0.28	0.52	−3.66	3.30
SERC-1	−0.86	1.16	−2.02	6.32	−0.37	0.99	−2.97	6.28	−0.31	1.66	−2.79	6.53
SERC-3	−0.06	1.49	−3.03	4.51	−0.42	1.50	−3.15	5.23	−1.15	3.23	−2.28	6.49
SERC-5	−0.32	2.02	−4.80	5.48	−0.52	2.40	−4.18	5.47	−0.85	3.16	−3.42	6.37

It should be noted that there are differences inherent in the way the photon-counting lidar systems operate as opposed to discrete-return lidars from which the DTMs/CHMs were derived; thus vegetation is probably sampled quite differently. The ICESat-2 SDT performed a comparative analysis between the full-resolution photon-counting data and the DTM/CHM (Neuenschwander et al., unpublished data). They found that at the Pine Barrens and SERC sites, the canopy height residuals include approximately 4 and 5 m average biases, respectively, with RMS errors ranging from 3.4 to 5.6 m. This could be partly explained by the difference in the way the canopy surface is delineated in these two truth data sets. Moreover, because the CHM was gridded into a 1 m product, slight differences in geo-location could have caused discrepancies. The interpolation itself could have resulted in a loss of information and resolution affecting the residuals. As a result the residuals we observed, when comparing the results to the DTMs/CHMs, are probably overestimated. Despite this limitation, in the absence of a ground truth data set, this comparison provides the only truly independent validation available.

Figures 6 and 7 show examples of retrieved ground and canopy surfaces *versus* the DTM and CHM, respectively.

7. Conclusion

The ATLAS instrument on-board NASA's ICESat-2 mission will use a technology newly applied to surface altimetry, with the main scientific objectives of measuring ice sheet elevation, sea ice freeboard (to enable thickness estimates), and canopy height (to enable vegetation biomass assessment). This new approach relies on using a high-repetition rate, low per-pulse laser energy (25–100 μJ), photon-counting detection mechanism. As the application of photon-counting altimetry from orbit has not previously been carried out, there are uncertainties about the data utility in regard to vegetation and ecosystem science objectives. The main challenge arises, generally, from the combination of transmitting low-energy laser pulses and the low reflectivity of soil and vegetation at the laser wavelength. Depending upon vegetation type, crown density, atmospheric conditions, and solar elevation, the number of detected signal photons reflected from the canopy-top and the underlying ground surface is expected to range between 0 and 10 for each laser shot. This, combined with solar background noise, will ultimately result in a very low SNR per footprint.

High-density foliage limits the amount of laser energy that penetrates through the canopy and reaches the surface. This results in low return from the ground, which

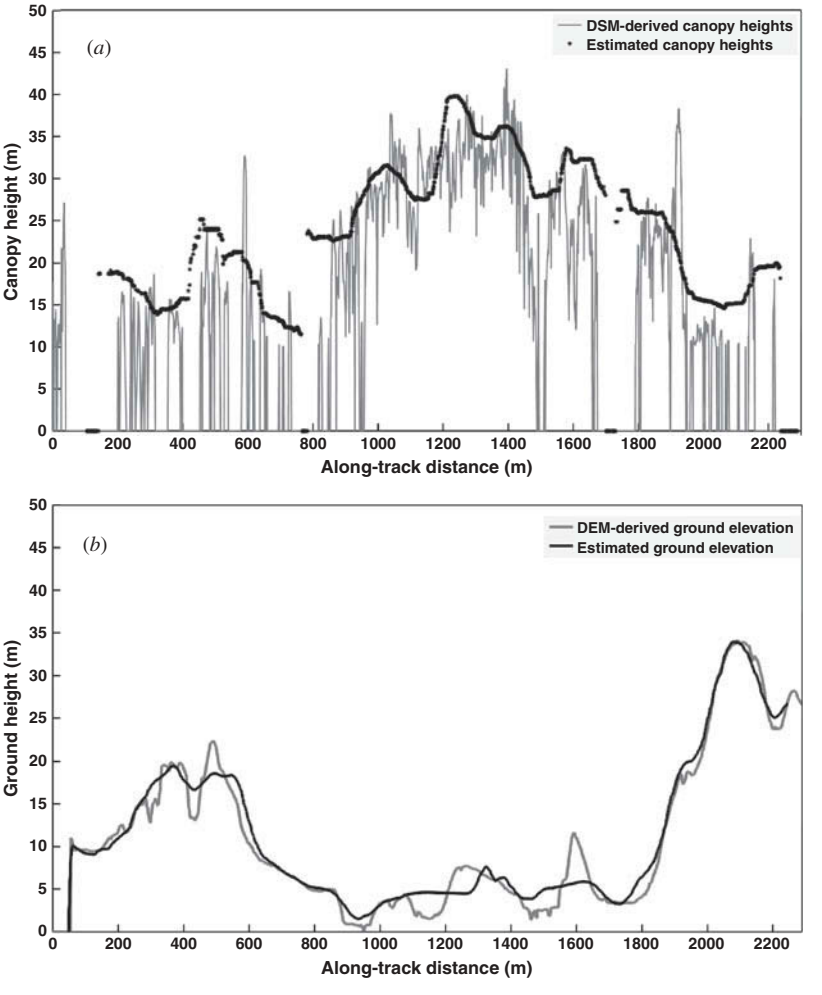

Figure 6. (*a*) Recovered canopy surface over the SERC-3 flight line (2 MHz noise rate) against the canopy height model (CHM). RMSE = 5.23 m; (*b*) Recovered ground surface over the SERC-3 flight line (2 MHz noise rate) against the digital terrain model (DTM). RMSE = 1.50 m.

subsequently makes canopy height retrieval challenging. On the other hand, low canopy closure might not provide a strong return signal from the canopy-top, which also affects canopy height estimation. Therefore, sophisticated techniques are required to accurately determine canopy height. In this article, an automatic technique for detecting ground and canopy-top surfaces from simulated ICESat-2 data is developed and validated. Data analysed in this study were created from airborne lidar campaigns by a Sigma Space Corporation photon-counting lidar and were sub-sampled to the photon density expected for ICESat-2. The signal extraction algorithm developed in this article is based on a multi-step mathematical and statistical process in which local neighbourhood information in the photon cloud is utilized. The algorithm uses statistical parameters from a histogram of elevations for 100 shots in the along-track direction to detect high-density photon clusters. A signal probability metric is then generated to facilitate extraction of ground and canopy

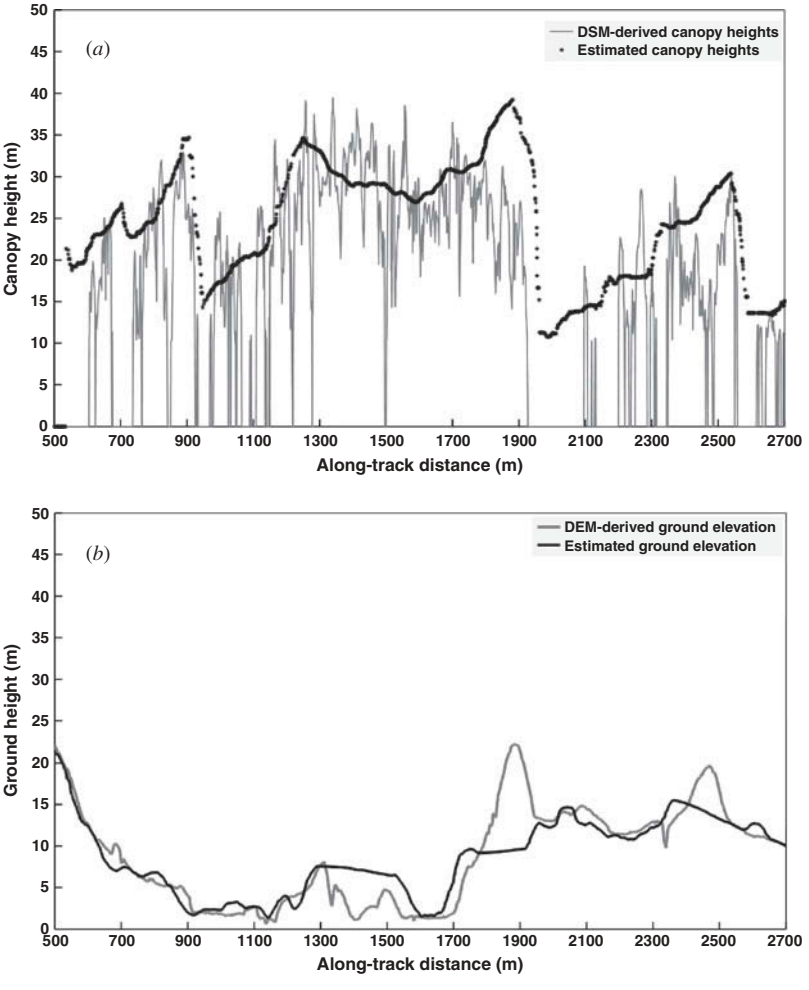

Figure 7. (*a*) Recovered canopy surface over the SERC-5 flight line (5 MHz noise rate) against the canopy height model (CHM). RMSE = 6.37 m; (*b*) Recovered ground surface over the SERC-5 flight line (5 MHz noise rate) against the digital terrain model (DTM). RMSE = 3.16 m.

cover surfaces. Subsequently, an iterative cubic spline interpolation between classified ground/canopy-top photons further refines the classification results and improves the detection accuracy.

Validation of results against two reference data sets showed that ground/canopy elevation can be estimated from simulated ICESat-2 data with reasonably high accuracy. At the Pine Barrens, ground surfaces were recovered with an average RMS of 0.3 m and canopy heights were resolved with an average accuracy of better than 3.0 m. However, canopy height retrievals over SERC forests, with significantly higher canopy cover, were more challenging due to penetration of fewer photons through the dense layers of canopy. At SERC, the average residuals for ground surface and canopy-tops were 1.2 and 5.0 m respectively.

The results presented here are for two temperate forests in the eastern USA, and thus the conclusion should not be generalized to other biomes such as the dense, multi-layered

canopy in tropical forests. Nonetheless, the algorithm performs reasonably well under a broad range of SNR scenarios. It should also be noted that regardless of the final ICESat-2 instrument parameters, the algorithm will still be applicable for detecting ground/canopy surfaces. Moreover, due to the algorithm's success in the challenging case of forested ecosystems, it is very likely to produce accurate results from photon-counting data collected over land ice and sea ice.

There are limits to the applicability of laser altimetry over densely vegetated areas. At some point, the laser energy extinction is too high for any detectable returns to be expected from the ground surface. This is especially the case for the ICESat-2 mission, since the instrument design is primarily driven by ice objectives. Despite this fact, our results suggest that ICESat-2 could provide valuable capabilities for determining forest canopy height; however, the degree of its success in this regard remains a topic of ongoing research.

Acknowledgements

This study was carried out in collaboration with the ICESat-2 project and Science Definition Team. We thank Bea Csatho, Thorsten Markus, Thomas Neumann, Benjamin Smith, and Ross Nelson (NASA – GSFC) for their support and helpful comments.

Funding

This work was supported by the National Aeronautics and Space Administration (NASA) [grant number NNX09AE54G].

References

Abdalati, W., H. J. Zwally, R. Bindschadler, B. Csatho, S. Farrell, H. Fricker, D. Harding, R. Kwok, M. Lefsky, T. Markus, A. Marshak, T. Neumann, S. Palm, R. Schutz, B. Smith, J. Spinhirn, C. Webb, and T. Neumann. 2010. "The ICESat-2 Laser Altimetry Mission." *Proceedings of the IEEE* 98 (5): 735–751. doi:10.1109/JPROC.2009.2034765.

Barbieri, K., A. Brenner, T. Markus, T. Neumann, J. Saba, and D. Yi. 2009. *Description of Subsampling Algorithm for Sigma Space Data*. Technical Report. Greenbelt, MD: NASA Goddard Space Flight Center.

Barbieri, K., A. Brenner, T. Markus, T. Neumann, J. Saba, D. Yi, and K. Brunt. 2010. *Description of ICESat-II Simulated Data Created from Sigma Space MPL Laser Data*. Technical Report. Greenbelt, MD: NASA Goddard Space Flight Center.

Collins, B. R., and K. H. Anderson. 1994. *Plant Communities of New Jersey: A Study in Landscape Diversity*. New Brunswick, NJ: Rutgers University Press.

Cossio, T. K., K. C. Slatton, W. E. Carter, K. Y. Shrestha, and D. Harding. 2010. "Predicting Small Target Detection Performance of Low-SNR Airborne Lidar." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3 (4): 672–688. doi:10.1109/JSTARS.2010.2053349.

Degnan, J. J. 2002. "Photon-Counting Multikilohertz Microlaser Altimeters for Airborne and Spaceborne Topographic Measurements." *Journal of Geodynamics* 34 (3–4): 503–549. doi:10.1016/S0264-3707(02)00045-5.

Fricker, H. A. 2005. "Assessment of ICESat Performance at the Salar De Uyuni, Bolivia." *Geophysical Research Letters* 32 (21): 3–7. doi:10.1029/2005GL023423.

Hancock, S., M. Disney, J. P. Muller, P. Lewis, and M. Foster. 2011. "A Threshold Insensitive Method for Locating the Forest Canopy Top with Waveform Lidar." *Remote Sensing of Environment* 115: 3286–3297. doi:10.1016/j.rse.2011.07.012.

Harding, D. J. 2005. "ICESat Waveform Measurements of Within-Footprint Topographic Relief and Vegetation Vertical Structure." *Geophysical Research Letters* 32 (21): 1–3. doi:10.1029/2005GL023471.

Herzfeld, U. C., B. W. Mcdonald, B. F. Wallin, T. A. Neumann, T. Markus, A. C. Brenner, and C. Field. 2013. "Algorithm for Detection of Ground and Canopy Cover in Micropulse Photon-Counting Lidar Altimeter Data in Preparation for the ICESat-2 Mission." *IEEE Transactions on Geoscience and Remote Sensing* 1: 1–1.

Kurtz, N. T., T. Markus, D. J. Cavalieri, W. Krabill, J. G. Sonntag, and J. Miller. 2008. "Comparison of ICESat Data with Airborne Laser Altimeter Measurements Over Arctic Sea Ice." *IEEE Transactions on Geoscience and Remote Sensing* 46 (7): 1913–1924. doi:10.1109/TGRS.2008.916639.

Kwok, R., G. F. Cunningham, H. J. Zwally, and D. Yi. 2007. "Ice, Cloud, and Land Elevation Satellite (ICESat) over Arctic Sea Ice: Retrieval of Freeboard." *Journal of Geophysical Research* 112 (C12): C12013. doi:10.1029/2006JC003978.

Lefsky, M. A., D. J. Harding, M. Keller, W. B. Cohen, C. C. Carabajal, F. Espirito-Santo, M. O. Hunter, and R. De Oliveira. 2005. "Estimates of Forest Canopy Height and Aboveground Biomass Using ICESat." *Geophysical Research Letters* 32 (22): L22S02. doi:10.1029/2005GL023971.

Liu, Z., W. Hunt, M. Vaughan, C. Hostetler, M. Mcgill, K. Powell, D. Winker, and Y. Hu. 2006. "Estimating Random Errors Due to Shot Noise in Backscatter Lidar Observations." *Applied Optics* 45 (18): 4437–4447. doi:10.1364/AO.45.004437.

Magruder, L. A., C. E. Webb, T. J. Urban, E. C. Silverberg, and B. E. Schutz. 2007. "ICESat Altimetry Data Product Verification at White Sands Space Harbor." *IEEE Transactions on Geoscience and Remote Sensing* 45 (1): 147–155. doi:10.1109/TGRS.2006.885070.

Martino, A. J. 2010. *ATLAS Performance Spreadsheet*. http://icesat.gsfc.nasa.gov/icesat2/data/sigma/sigma_data.php.

Morsdorf, F., O. Frey, E. Meier, K. I. Itten, and B. Allgöwer. 2008. "Assessment of the Influence of Flying Altitude and Scan Angle on Biophysical Vegetation Products Derived from Airborne Laser Scanning." *International Journal of Remote Sensing* 29 (5): 1387–1406. doi:10.1080/01431160701736349.

Pang, Y., M. A. Lefsky, H. E. Andersen, M. E. Miller, and K. Sherrill. 2008. "Validation of the ICESat Vegetation Product Using Crown-Area-Weighted Mean Height Derived Using Crown Delineation with Discrete Return Lidar Data." *Canadian Journal of Remote Sensing* 34 (S2): S471–S484. doi:10.5589/m08-074.

Parker, G. 2012. *Smithsonian Environmental Research Center, Edgewater, MD, USA*. http://www.ctfs.si.edu/site/SERC%3A+Smithsonian+Environmental+Research+Center.

Popescu, S. C., K. Zhao, A. Neuenschwander, and C. Lin. 2011. "Satellite Lidar Vs. Small Footprint Airborne Lidar: Comparing the Accuracy of Aboveground Biomass Estimates and Forest Structure Metrics at Footprint Level." *Remote Sensing of Environment* 115 (11): 2786–2797. doi:10.1016/j.rse.2011.01.026.

Rosette, J. A., P. R. J. North, and J. C. Suárez. 2008. "Vegetation Height Estimates for a Mixed Temperate Forest Using Satellite Laser Altimetry." *International Journal of Remote Sensing* 29 (5): 1475–1493. doi:10.1080/01431160701736380.

Schutz, B. E., H. J. Zwally, C. Shuman, D. Hancock, and J. P. Dimarzio. 2005. "Overview of the ICESat Mission." *Geophysical Research Letters* 32 (21): 1–4. doi:10.1029/2005GL024009.

Shuman, C., H. J. Zwally, B. E. Schutz, A. C. Brenner, J. P. Dimarzio, V. P. Suchdeo, and H. A. Fricker. 2006. "ICESat Antarctic Elevation Data: Preliminary Precision and Accuracy Assessment." *Geophysical Research Letters* 33 (7): 10–13. doi:10.1029/2005GL025227.

Thomas, R., C. Davis, E. Frederick, W. Krabill, Y. Li, S. Manizade, and C. Martin. 2008. "A Comparison of Greenland Ice-Sheet Volume Changes Derived from Altimetry Measurements." *Journal of Glaciology* 54 (185): 203–212. doi:10.3189/002214308784886225.

Yang, Y., A. Marshak, S. P. Palm, T. Varnai, and W. J. Wiscombe. 2011. "Cloud Impact on Surface Altimetry from a Spaceborne 532-Nm Micropulse Photon-Counting Lidar: System Modeling for Cloudy and Clear Atmospheres." *IEEE Transactions on Geoscience and Remote Sensing* 49 (12): 4910–4919. doi:10.1109/TGRS.2011.2153860.

Yi, D., H. J. Zwally, and X. Sun. 2005. "ICESat Measurement of Greenland Ice Sheet Surface Slope and Roughness." *Annals of Glaciology* 42 (1): 83–89. doi:10.3189/172756405781812691.

Yu, A. W., M. A. Stephen, S. X. Li, G. B. Shaw, A. Seas, E. Dowdye, E. Troupaki, P. Poulios, K. Mascetti, K. Mascetti, W. A. Clarkson, N. Hodgson, and R. K. Shori. 2010. "Space Laser Transmitter Development for ICESat-2 Mission." *Solid State Lasers XIX: Technology and Devices* 7578: 757809-1–757809-11. doi:10.1117/12.843342.

Zimble, D. A., D. L. Evans, G. C. Carlson, R. C. Parker, C. G. Grado, and P. D. Gerard. 2003. "Characterizing Vertical Forest Structure Using Small-Footprint Airborne Lidar." *Remote Sensing of Environment* 87: 171–182. doi:10.1016/S0034-4257(03)00139-1.

Zwally, H. J., M. B. Giovinetto, J. Li, H. G. Cornejo, M. A. Beckley, A. C. Brenner, J. L. Saba, and D. Yi. 2005. "Mass Changes of the Greenland and Antarctic Ice Sheets and Shelves and Contributions to Sea-Level Rise: 1992–2002." *Journal of Glaciology* 51 (175): 509–527. doi:10.3189/172756505781829007.

Zwally, H. J., D. Yi, R. Kwok, and Y. Zhao. 2008. "ICESat Measurements of Sea Ice Freeboard and Estimates of Sea Ice Thickness in the Weddell Sea." *Journal of Geophysical Research* 113 (C2): C02S15. doi:10.1029/2007JC004284.